

Learning Structured Knowledge from Social Tagging Data:

a critical review of methods and techniques

The 8th IEEE International Conference on Social Computing and
Networking (SocialCom 2015), Chengdu, China, 19-21 Dec 2015

Hang Dong, Wei Wang, Hai-Ning Liang
CSSE, Xi'an Jiaotong-Liverpool University



Xi'an Jiaotong-Liverpool University

西交利物浦大學

Email: Hang.Dong@xjtlu.edu.cn

Last edited on 12th Jan 2016 by Hang Dong.

Content

- Background: social tagging & Folksonomies
- Knowledge Structures / Ontologies
- Methods & techniques for learning KS from social tagging data
- Issues & future studies

Background: Folksonomies

- **Folksonomy = Folk + Taxonomy** (by Thomas Vander Wal in 2007, <http://www.vanderwal.net/folksonomy.html>)
- Share a resource – create personal free short text description
- Social Tagging Systems: delicious, flickr, douban, bibsonomy...
- Noisy, ambiguous then not very useful.

Noisy & Complex Social Tags

- apple (polysemy)
- to-read, myThesis (personal use)
- sematnikweb (typo, multiword)
- data-mining, DataMining (multiword)
- \$\$\$\$Tsk (nonsense tag)
- Mac,Macintosh,Apple (synonym)
- Programming, Javascript, perl (gen-spec)

From Social Tagging Data to Ontologies

Entity Linking with a Knowledge Base: Issu... 3

Wei Shen, Jianyong Wang, and Jiawei Han. *Transactions on Knowledge & Data Engineering* 27(2):443--460 (2015)

10 hours and 2 minutes ago by @jaeschke

background base entity knowledge linking ner

★★★★★ (0)



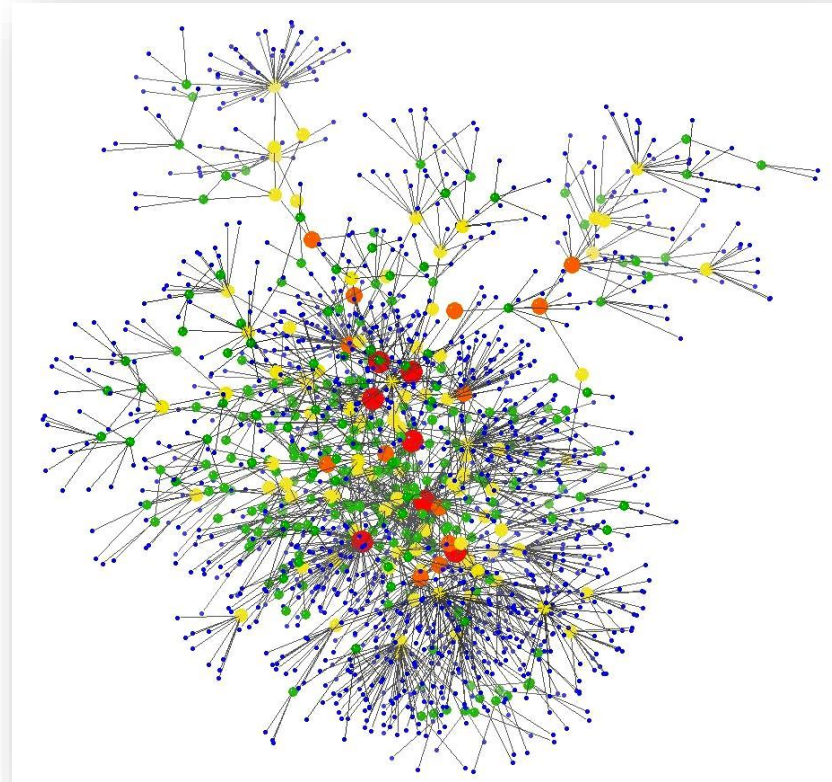
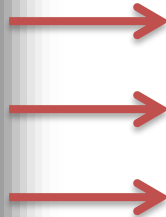
Knowledge-based systems: special issue o... 2

Khaldoun Zreik, and Cherif Branki. *Knowl.-Based Syst.* 13(1):1 (2000)

16 hours and 5 minutes ago by @chatelp

CyberDesign knowledge

★★★★★ (0)



<http://www.bibsonomy.org/tag/knowledge>

**Researcher generated data
(user-tag-resource-date)**

Acknowledgement to the figure from M. Triana. *Ontology... what?* (Jan 2012)

<http://www.micheltriana.com/blog/2012/01/20/ontology-what>

Useful and evolving ontologies

The structure of social tagging data

A. GARCIA-SILVA ET AL.

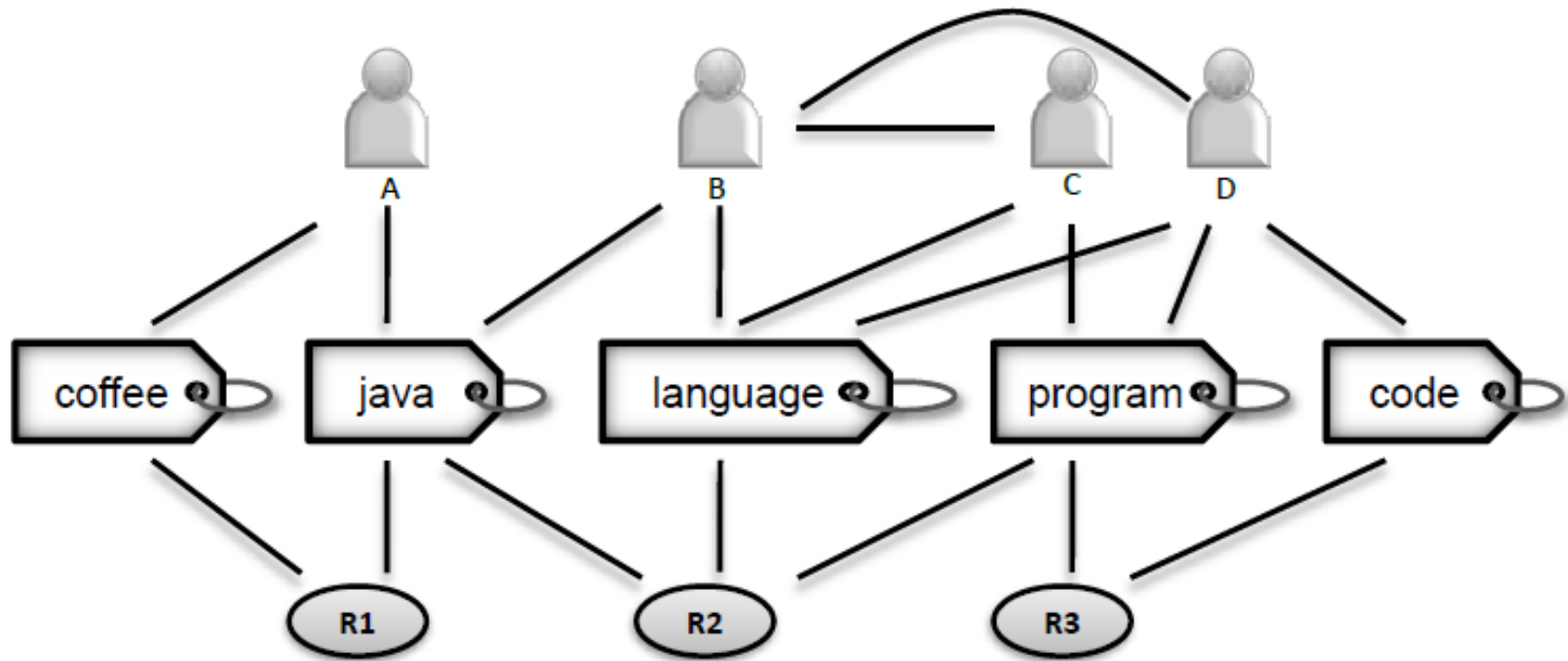


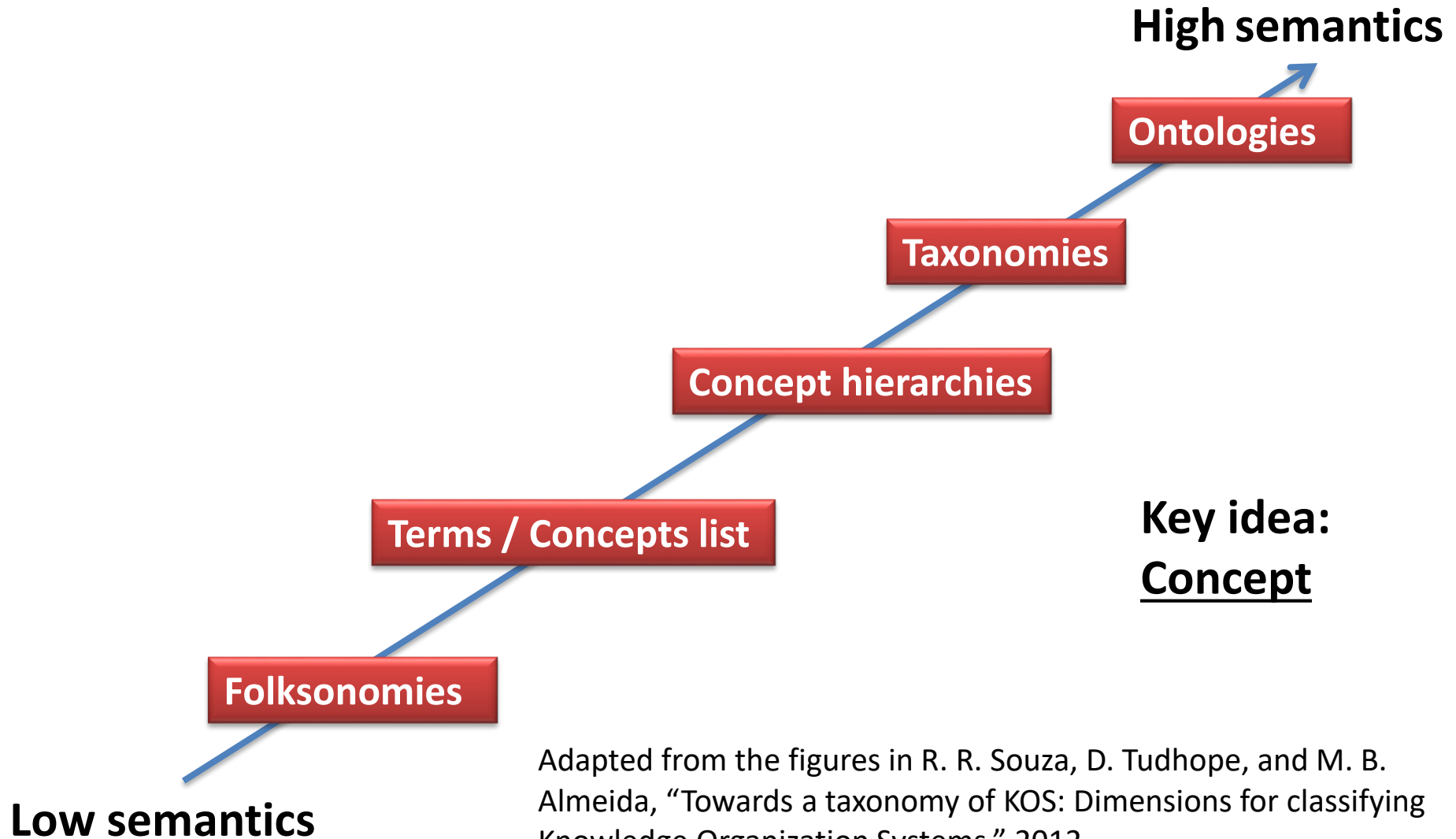
Figure 3 Graphical representation of the folksonomy example

Acknowledgement to the figure from A. García-Silva, O. Corcho, H. Alani, and A. Gómez-Pérez, "Review of the state of the art: Discovering and associating semantics to tags in folksonomies," *The Knowledge Engineering Review*, vol. 27, no. 01, pp. 57-85, Mar. 2012.

Applications

- General
 - (Semantic-based) Recommender systems
 - Ontology enrichment
 - Information retrieval and navigation
- Specific
 - E-learning discovery
 - Enterprise search and knowledge management
 - Academic communication
 - Image annotation, recognition, representation
 - Music mood mining

A spectrum of Knowledge Structures / ontologies



Adapted from the figures in R. R. Souza, D. Tudhope, and M. B. Almeida, "Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems," 2012.

Ontology learning layer cake

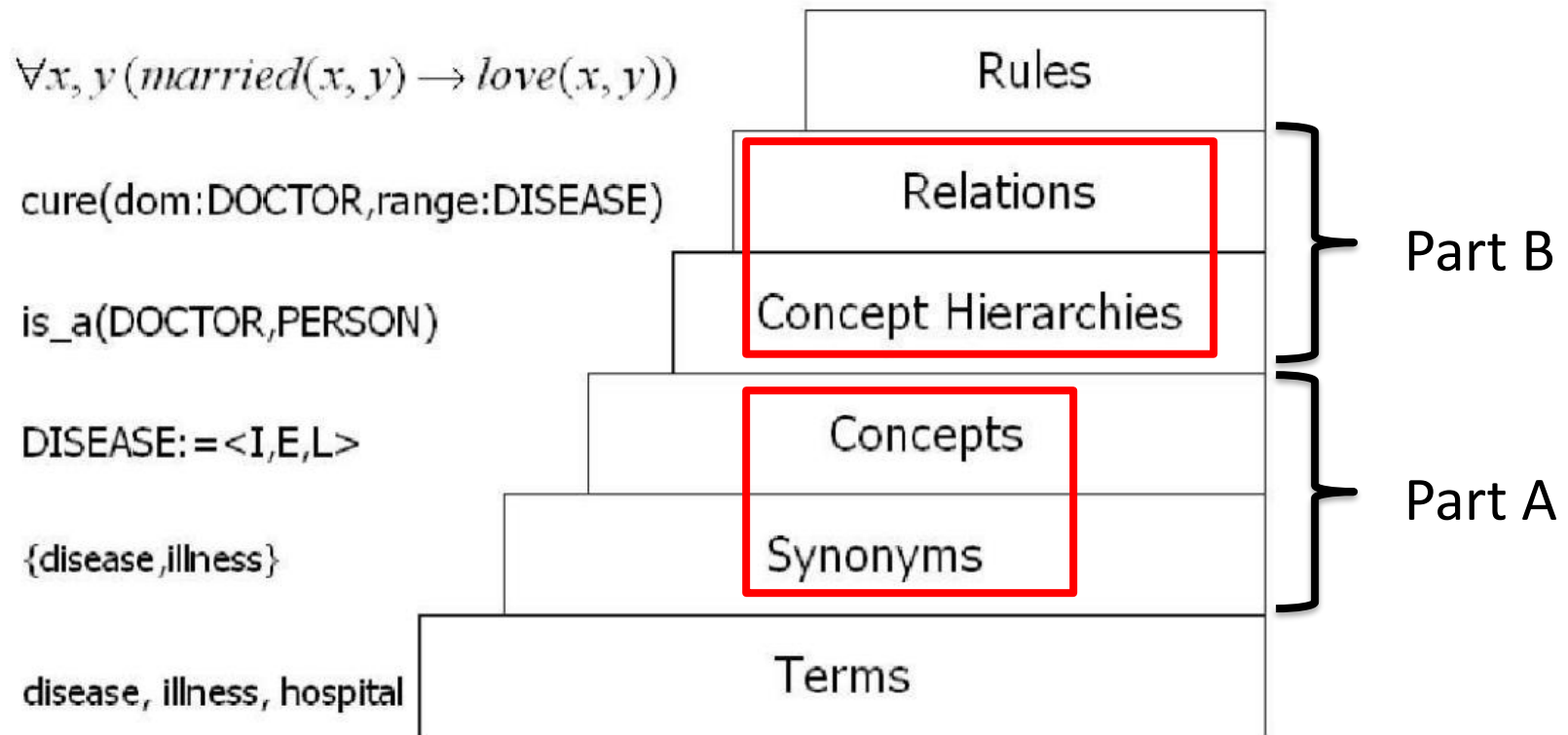
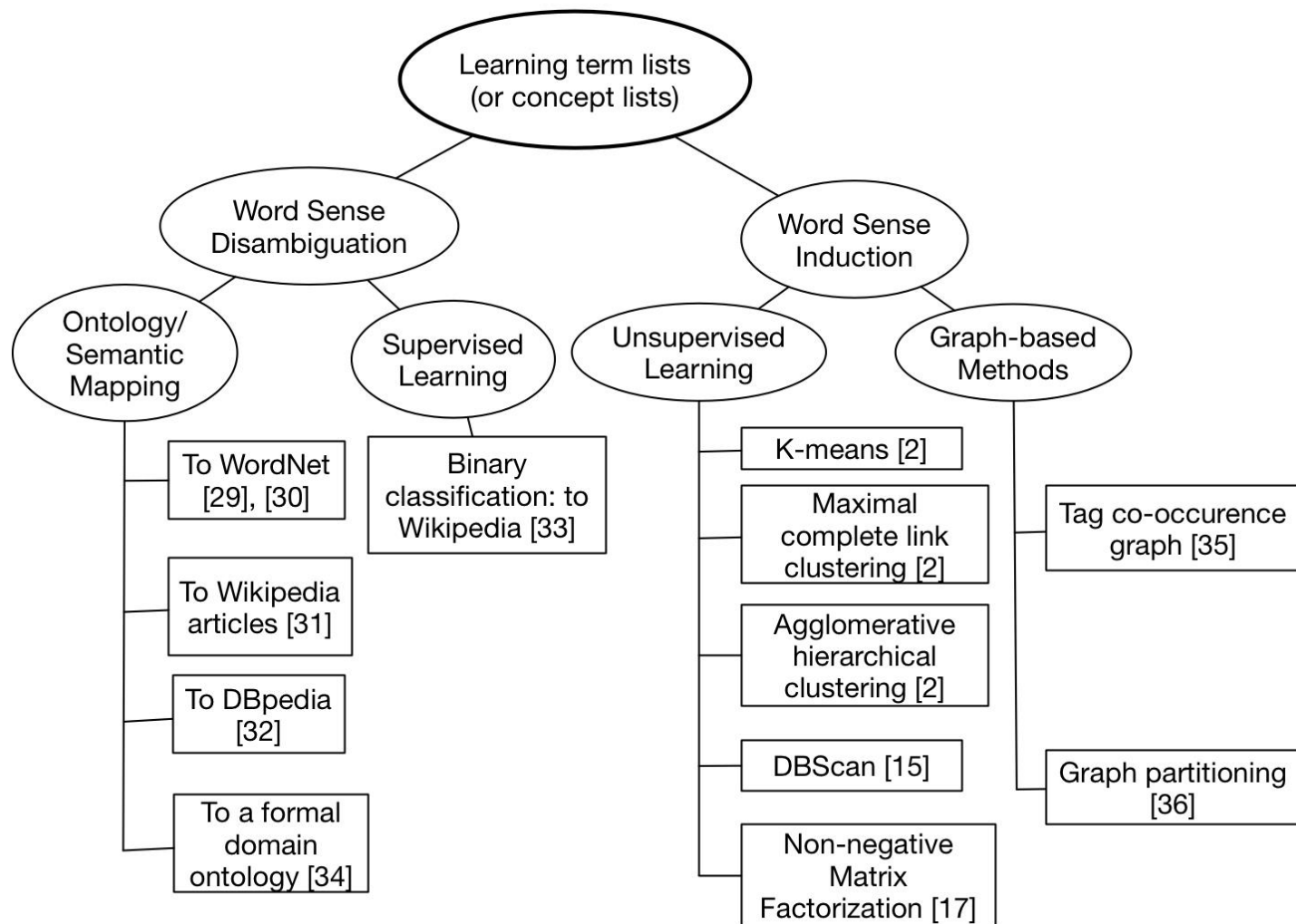


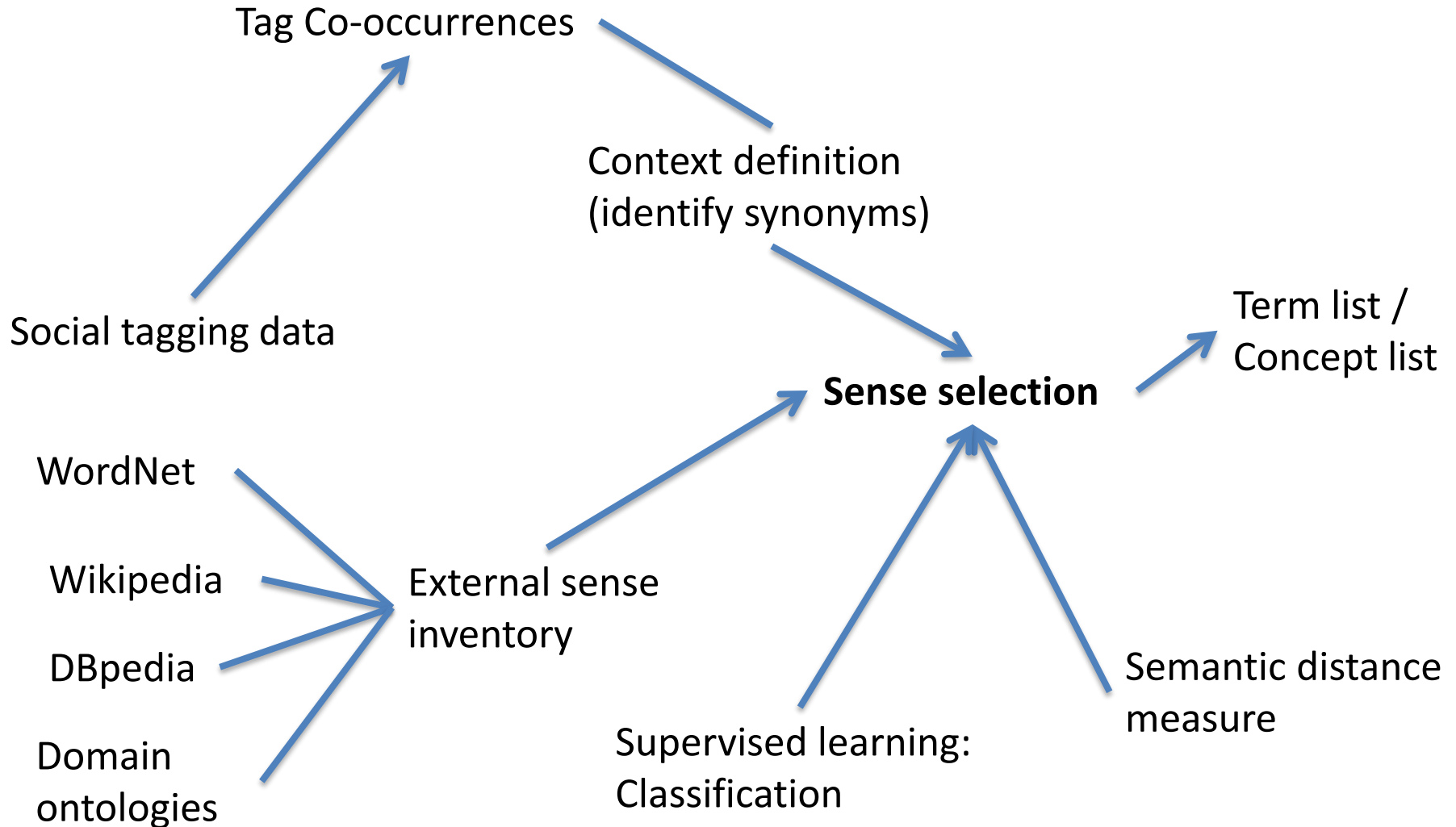
Figure 1. Ontology Learning Layer Cake

Adapted from the Figure 1 in Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini: 'Ontology Learning from Text: An Overview', 2003

Learning term (or concept) lists

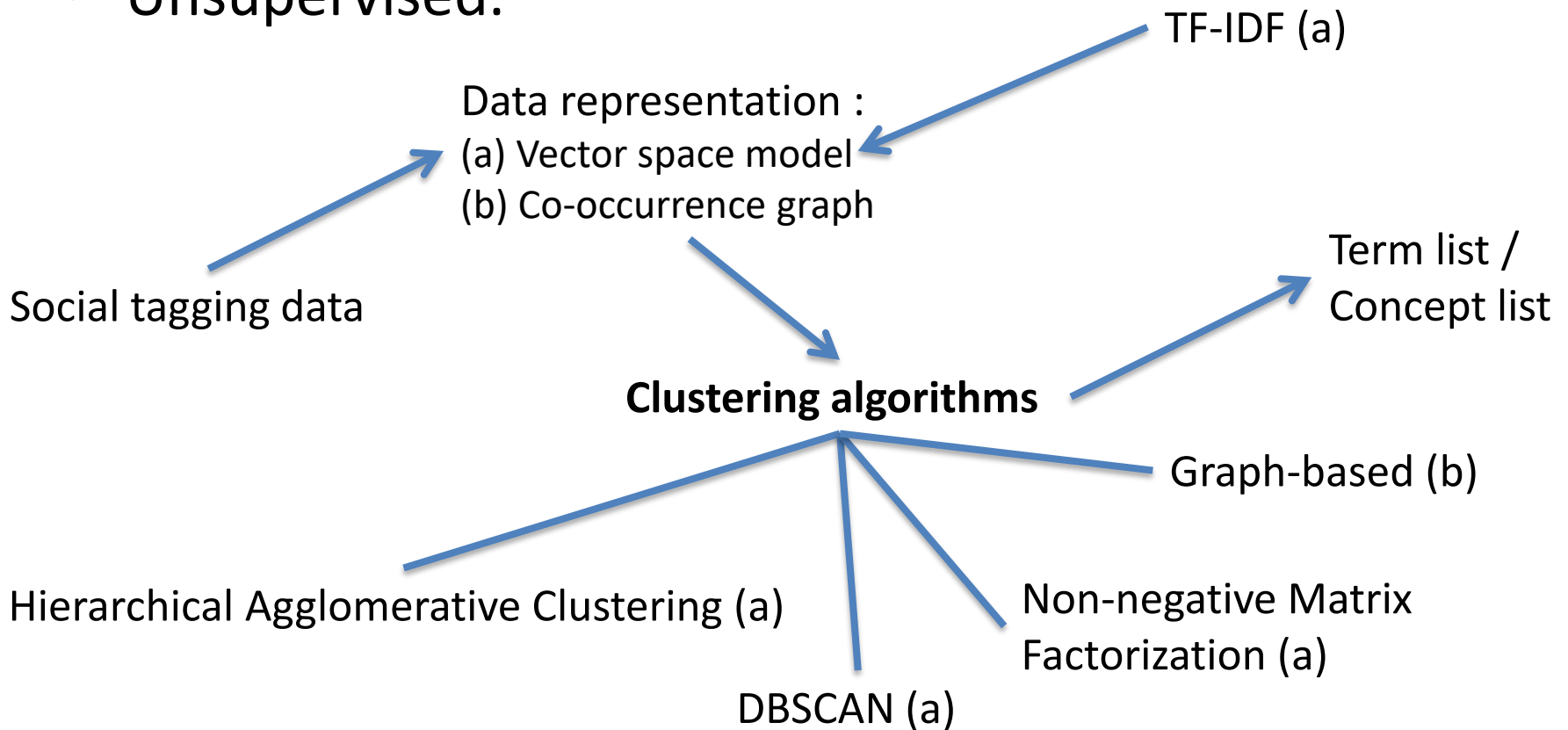


Word Sense Disambiguation

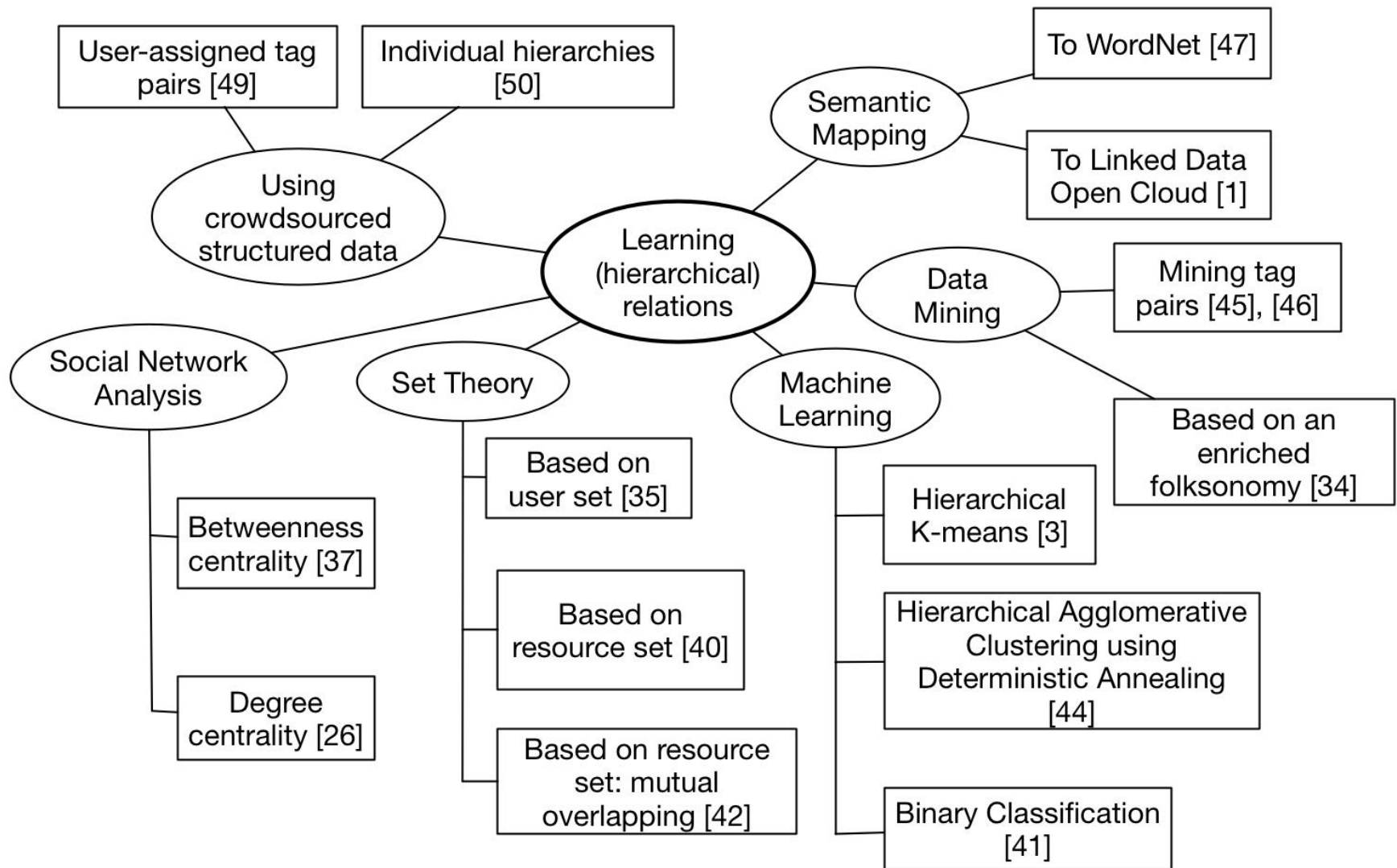


Word Sense Induction

- No external sense inventory.
- Unsupervised.



Learning (hierarchical) relations



“popularity-generalality” assumption

- The more popular/influential/frequent a tag is, the more general it is! [51]

If $\text{sim}(\text{tag1}, \text{tag2}) \geq \text{threshold}$ then do

 If tag1 is more popular than tag2 then do

 add tag2 as the child of tag1 [37]

Popularity metrics: Set Theory

- tag sets, user sets and/or resources sets

Algorithm 1 [35]:

If $R(\text{tagB}) \subseteq R(\text{tagA})$ then do
 add tagB as a child of tagA

Algorithm 2 [40]:

$$\begin{aligned} \text{inc}(\text{tagA}, \text{tagB}) &= |R(\text{tagA}) \cap R(\text{tagB})| / |R(\text{tagA})| \\ &= \text{gen}(\text{tagB}, \text{tagA}) \end{aligned}$$

Popularity metrics: Centrality in SNA

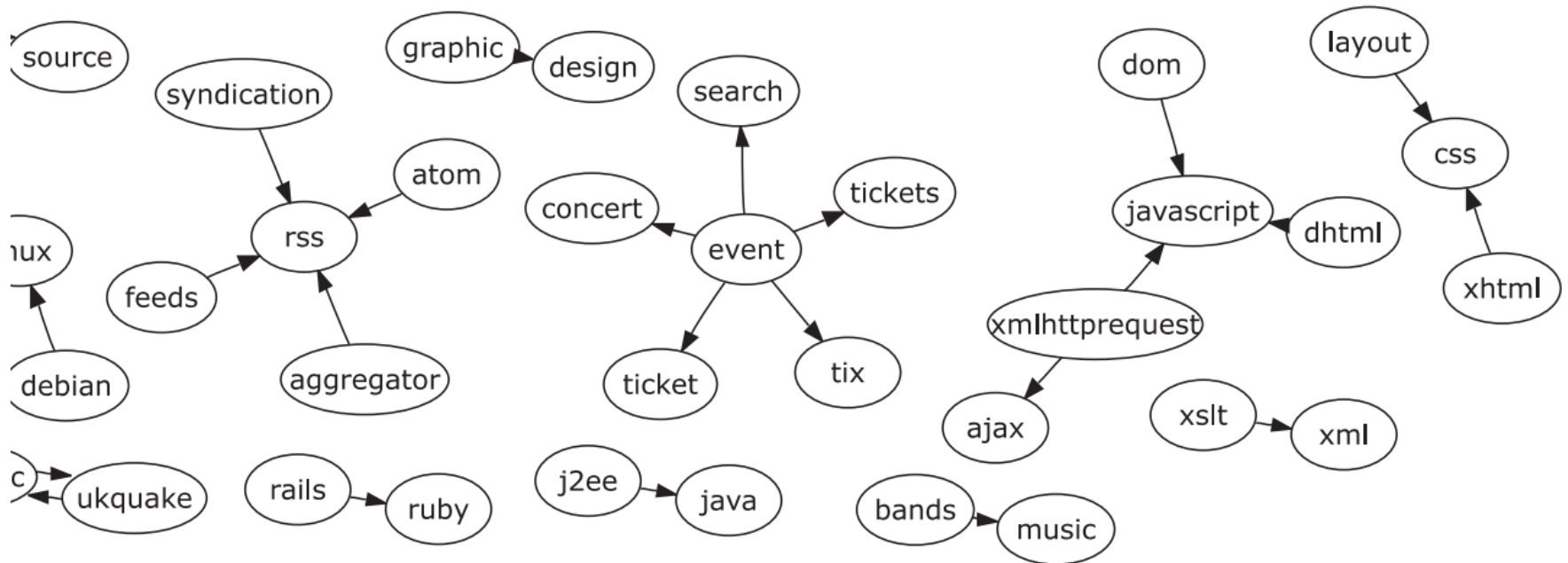
Social Network Analysis (SNA)

- Build a graph: nodes are tags, edges are their similarities [37]
- Measure how popular/influential a tag is:
 - Degree centrality
 - Betweenness centrality

Association rule mining

- Support and Confidence [45] [41]
- Important issue: Aggregate the tripartite structure into two-dimensional space.
- Association analysis is not enough to be used solely for hierarchical relation extraction.

Association rule mining



minsupp = .05%, mincon = 50%

Acknowledgement to the figure from C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme, "Mining Association Rules in Folksonomies," *Data Science and Classification, Studies in Classification, Data Analysis, and Knowledge Organization V. Batagelj, H.-H. Bock, A. Ferligoj and A. Žiberna, eds., pp. 261-270: Springer Berlin Heidelberg, 2006.*

Supervised learning

- Binary classification: hierarchical or not [41]
- Feature extraction:
 - Support and confidence
 - Cosine similarity
 - Inclusion and generalization measures
 - ...
- Instance labeling using WordNet and ConceptNet
- Class Imbalance Problem
- C 4.5, Naïve Bayes, Random Forest, AdaBoost, Log. Regression, SVM

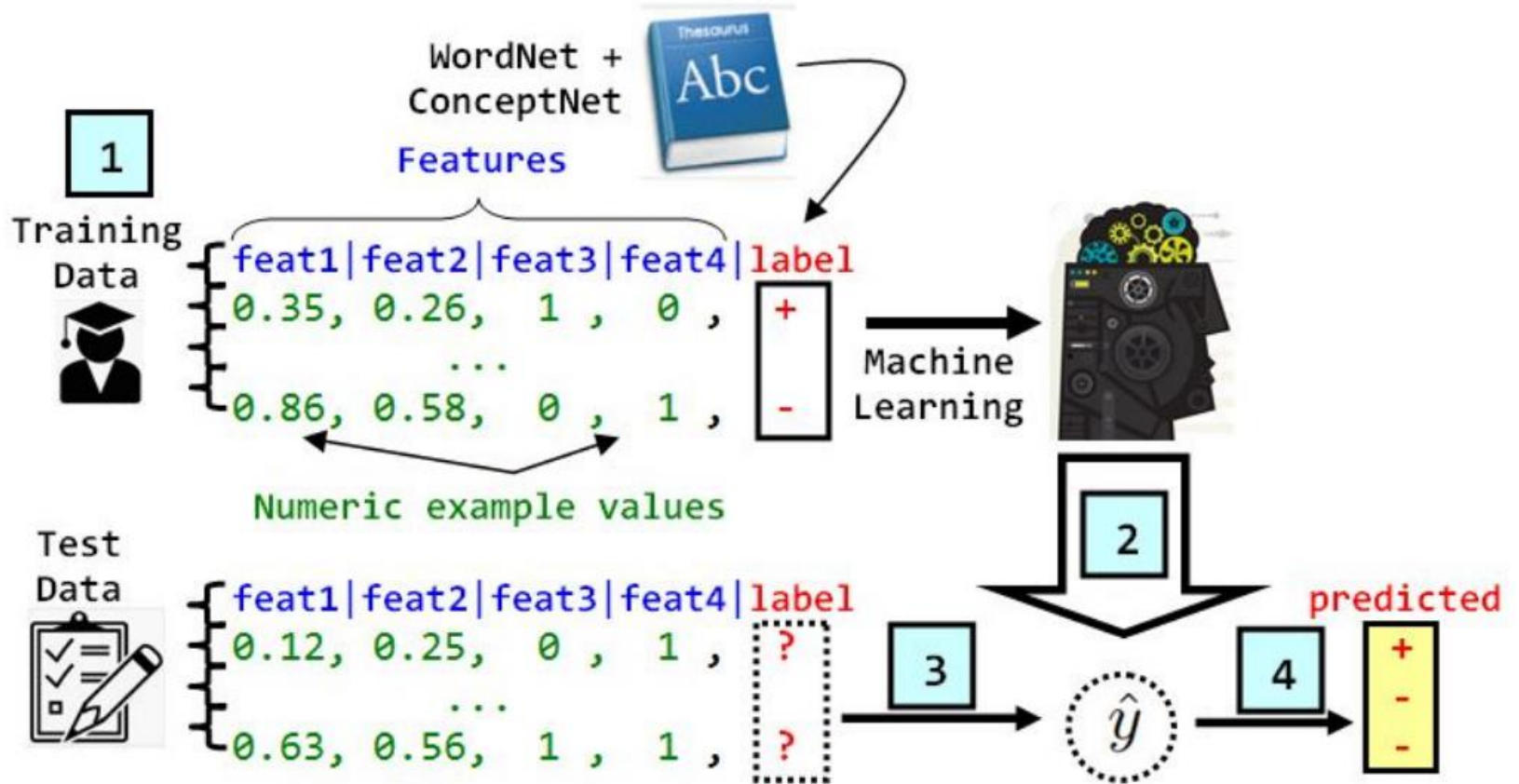


Figure 1: Classification for detecting subsumption relations between tags.

Acknowledgement to the figure from A. S. C. Rêgo, L. B. Marinho, and C. E. S. Pires, "A supervised learning approach to detect subsumption relations between tags in folksonomies," in the Proc. 30th Annu. ACM Symp. Applied Computing (SAC '15), Salamanca, Spain, 2015, pp. 409-415.

Crowdsourcing & semi-supervised methods

- Special functions of social tagging systems
 - Allowing users to annotate with tag pairs or hierarchies (TagTree, Delicious, Bibsonomy).
 - Some implicit functions: Flickr's photo management: photo-set-collection [50].
- Using relational clustering algorithms to aggregate the hierarchies [50].

Issues of existing studies

- External resources - not suitable for social tagging data.
- Only co-occurrence features – not precise.
- Unsupervised learning – need to specify # of clusters, can generate odd clusters.
- The “popularity-generality” assumption – not accurate.
- Crowdsourcing methods – not natural, more work for users, few data.
- Not showing the evolution of knowledge.

Future studies

- Using integral or new methods to derive semantics from social tagging data
- Investigate the evolution of knowledge in Folksonomies
 - From User-Resource-Tag to User-Resource-Tag-Date
- Big Data + ontology learning
- Specific domains + applications
 - Image
 - Academic resources
 - Music

Key References

- [2] J. Gemmell, A. Shepitsen, M. Mobasher, and R. Burke, “Personalization in folksonomies based on tag clustering,” in *the Proc. 6th Workshop Intelligent Techniques for Web Personalization and Recommender Systems*, Chicago, Illinois, USA, 2008.
- [15] P. Andrews and J. Pane, “Sense induction in folksonomies: a review,” *Artificial Intelligence Review*, vol. 40, no. 2, pp. 147-174, Aug. 2013.
- [18] A. García-Silva, O. Corcho, H. Alani, and A. Gómez-Pérez, “Review of the state of the art: Discovering and associating semantics to tags in folksonomies,” *The Knowledge Engineering Review*, vol. 27, no. 01, pp. 57-85, Mar. 2012.
- [30] S. Angeletou, M. Sabou, and E. Motta, ‘Semantically enriching folksonomies with FLOR,’ in *Proc. 1st Int. Workshop Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008) at The 5th Annu. European Semantic Web Conf. (ESWC 2008)*, Tenerife, Spain, 2008.
- [33] A. Joorabchi, M. English, and A. E. Mahdi, “Automatic mapping of user tags to Wikipedia concepts: The case of a Q&A website – StackOverflow,” *Journal of Information Science*, published online before print, May 2015.
- [35] P. Mika, “Ontologies are us: A unified model of social networks and semantics,” in *The Semantic Web–ISWC 2005*. Springer, 2005, pp. 522-536.
- [37] P. Heymann and H. Garcia-Molina. (2006, Apr. 26). *Collaborative creation of communal hierarchical taxonomies in social tagging systems* [online]. Technical Report, Stanford InfoLab. Available: <http://ilpubs.stanford.edu:8090/775/>

Key References

- [40] P. De Meo, G. Quattrone, and D. Ursino, “Exploitation of semantic relationships and hierarchical data structures to support a user in his annotation and browsing activities in folksonomies,” *Information Systems*, vol. 34, no. 6, pp. 511-535, Sep. 2009.
- [41] A. S. C. Rêgo, L. B. Marinho, and C. E. S. Pires, “A supervised learning approach to detect subsumption relations between tags in folksonomies,” in *the Proc. 30th Annu. ACM Symp. Applied Computing (SAC '15)*, Salamanca, Spain, 2015, pp. 409-415.
- [45] C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme, “Mining association rules in folksonomies,” in *Data Science and Classification*, V. Batagelj, H.-H. Bock, A. Ferligoj, & A. Žiberna, Eds. Springer Berlin Heidelberg, 2006, pp. 261-270.
- [50] A. Plangprasopchok, K. Lerman, and L. Getoor, “Growing a tree in the forest: constructing folksonomies by integrating structured metadata,” in *the Proc. of the 16th ACM SIGKDD Int. Conf. Knowledge discovery and data mining*, Washington, DC, USA, 2010.
- [51] F. Almoqhim, D. Millard, and N. Shadbolt, “Improving on popularity as a proxy for generality when building tag hierarchies from folksonomies,” in *Social Informatics*, Vol. 8851, L. Aiello and D. McFarland, Eds. Springer International Publishing, 2014, pp. 95-111.

The full list of references can be found in the paper:

Hang Dong, Wei Wang, Hai-Ning Liang. “*Learning Structured Knowledge from Social Tagging Data: A Critical Review of Methods and Techniques*,” in The 8th IEEE International Conference on Social Computing and Networking (SocialCom 2015). (accepted)

Q & A

Thank you for your attention